UNITED STATES DEPARTMENT OF AGRICULTURE
BUREAU OF AGRICULTURAL ECONOMICS

------0------

THE USE OF PUNCHED CARD TABULATING EQUIPMENT IN MULTIPLE CORRELATION PROBLEMS

Collected and prepared for the use of
Statisticians of the Bureau.

by

Bradford B. Smith, In Charge
Machine Tabulation and Computing Section

Washington, D. C.
October, 1923.

The use of punched card tabulating equipment is not new: It was used a number of years ago in the Bureau of the Census for the preparation of Correlation tables. More recently it has been used in at least two of the Bureaus of the Department of Agriculture in a similar way. The writer in no sense claims to be the sole author of the methods herein set forth except insofar as the special technique is devised to suit the equipment and problems of the Bureau of Agricultural Economics.

Especial credit is due to Messrs. Tolley and Ezekiel of this Bureau for devising the least square method (page 15 et seq.) of approaching correlation problems. The writer also wishes to express his appreciation to them for going over the following text, their suggestions and helpful criticism.

The coefficient of Multiple Correlation, R, is a measure of the degree of agreement between a given series and the estimated (generally forecasted) values of the same series, when these estimated values are determined mathematically from two or more different (independent) series.* The following pages describe the forms and arithmetic used in computing the value of R and in deriving the "regression equation" or forecasting formula. A five variable (four independent) example is given.

---

*The coefficient of Multiple Correlation (R) may be defined as the coefficient of correlation (Pearsonian) between the dependent variable and corresponding estimates thereof, computed from the independent variables, i. e. if X', the estimated value of X, equals b_1A plus b_2B and b_3C......plus a constant, where A, B, and C, are ~~the independent variables~~ and b_1, b_2, b_3, are the net regression coefficients, R equals the mean product of the deviations of X and X' from their respective averages divided by the product of the standard deviations of X and X'. R is also equal to the standard deviation of X' divided by the standard deviation of X. Also, the standard deviation of X' squared is equal to the mean product of the deviations of X and X' from their respective averages.

CODING The data should be coded so that the variation in each series is somewhere between ten and thirty, depending upon the discretion of the person conducting the investigation. The smaller the variation, the simpler is the process of performing the calculations, for the whole success of using Hollerith Machines* lies in the grouping of like values.** On the other hand, too great a grouping--or reduction of variation--results in reduced accuracy. The coding may be accomplished either by subtraction or by division or both: Thus if the series under consideration varies between 120 and 175, subtract 120 from all items, and divide the differences by 5 thus reducing the variation to from 0 to 11. This coding of the series in no way invalidates the calculation of R. Coding by division slightly reduces accuracy, but subtracting in no way affects accuracy.

On the following pages are given the coded series used in the example. A, B, C, & D are the independent variables. X is the dependent variable (the values of which it is desired to estimate, or "predict".) The series should be so arranged that the dependent variable is placed in a column to the right of the independents. After the coded series have been listed, cross-add the values for A, B, C, and D, & X for each observation and list the sum in the "Check Sum" column on the right, Col. S in the tabulation. This check sum column is not essential to the solution but is of considerable aid in checking the arithmetic.

CHECK SUM

---

* There are several types of punched card sorting and tabulating machines. Those used in this Bureau are for the most part Hollerith Machines.

** Moreover, if the extensions are to be made by hand, figures that can be readily multiplied mentally greatly facilitate the process.

DATA FOR EXAMPLE

Record No.	$A^{(1)}$	$B^{(2)}$	$C^{(3)}$	$D^{(4)}$	$X^{(5)}$	S
1	10	4	7	00	10	31
2	3	10	11	3	11	38
3	10	5	12	00	6	33
4	10	3	11	2	10	36
5	10	6	10	3	5	34
6	6	9	1	7	10	33
7	3	3	10	4	0	20
8	0	7	5	1	5	18
9	3	6	10	8	1	28
10	3	1	12	5	4	25
11	6	6	8	0	6	26
12	10	6	2	4	6	28
13	3	5	2	6	4	20
14	12	3	5	0	2	22
15	12	6	0	12	2	32
16	12	4	5	4	6	31
17	6	3	11	3	3	26
18	10	12	8	11	11	52
19	3	6	8	8	3	28
20	10	3	4	3	5	25
21	10	3	11	12	12	48
22	10	3	10	3	7	33
23	3	5	00	10	4	22
24	6	3	8	00	7	24
25	0	1	11	7	5	24
26	10	5	8	0	5	28
27	10	4	0	9	4	27
28	10	6	12	6	4	38
29	12	3	10	12	1	38
30	10	5	12	2	0	29
31	12	9	7	1	4	33
32	10	5	11	3	2	31
33	7	6	4	4	11	32
34	6	4	10	0	12	32
35	7	4	12	6	12	41
36	3	5	0	10	8	26
37	3	5	5	4	2	19
38	3	1	0	12	4	20
39	3	0	1	8	1	13
40	12	3	11	9	4	39
41	10	1	8	0	12	31
42	10	3	10	1	12	36
43	6	5	5	0	0	16
44	10	3	11	1	5	30
45	12	5	5	4	12	38
46	7	5	5	1	3	21
47	12	7	10	0	0	29
48	10	2	7	3	4	26
49	10	3	7	0	6	26
50	10	3	4	4	11	32
51	12	4	5	1	2	24
52	3	4	4	3	4	18
53	10	4	5	3	0	22
54	3	9	10	0	1	23
55	6	2	10	8	1	27

Record No.	A	B	C	D	X	S
56	6	3	7	1	5	22
57	10	3	11	2	11	37
58	6	6	5	0	3	20
59	10	3	8	0	12	33
60	12	3	4	4	0	23
61	10	7	5	0	2	24
62	10	1	5	3	1	20
63	6	6	8	1	1	22
64	12	3	10	0	12	37
65	10	1	4	4	7	26
66	12	6	5	1	10	34
67	3	2	5	3	0	13
68	12	8	11	11	12	54
69	6	2	10	7	2	27
70	12	6	11	1	3	33
71	10	7	8	0	5	30
72	12	4	7	3	2	28
73	10	3	10	4	3	30
74	12	4	8	0	0	24
75	12	5	7	3	12	39
76	10	1	7	1	7	26
77	3	6	8	6	2	25
78	6	5	11	5	0	27
79	12	6	8	8	2	36
80	3	3	11	2	8	27
81	10	6	7	6	1	30
82	10	6	4	1	6	27
83	6	8	8	5	3	30
84	10	3	11	7	2	33
85	10	4	2	7	5	28
86	10	4	11	1	1	27
87	6	3	5	1	6	21
88	12	6	8	0	0	26
89	12	2	0	12	12	38
90	6	2	8	11	12	39
91	6	2	0	12	8	28
92	10	4	2	6	2	24
93	3	4	10	3	2	22
94	6	3	1	7	2	19
95	3	1	7	1	8	20
96	10	3	8	8	12	41
97	6	0	0	4	11	21
98	10	6	5	0	4	25
99	3	7	7	3	4	24
100	7	4	8	12	0	31
101	6	5	5	3	4	23
102	10	3	5	1	2	21
103	0	3	5	0	3	11
104	0	2	0	6	2	10
105	0	4	0	1	2	7
106	3	6	0	7	6	22
107	10	1	10	1	0	22
108	10	6	4	3	2	25
109	10	4	7	0	6	27

Record No.	A	B	C	D	X	S
110	0	1	0	3	0	4
111	0	5	2	3	0	10
112	10	0	11	2	2	25
113	6	4	0	7	0	17
114	0	5	7	1	3	16
115	3	3	8	0	7	21
116	10	5	11	1	0	27
117	10	1	0	7	2	20
118	10	2	5	3	5	25
119	0	6	12	8	9	35
120	6	3	11	1	9	30
121	12	3	5	4	0	24
122	10	5	11	3	0	29
123	10	3	8	0	4	25
124	10	4	5	3	4	26
125	10	6	11	10	1	38
126	3	6	11	4	4	28
127	10	6	10	12	2	40
128	6	4	0	7	2	19
129	10	3	11	3	7	34
130	3	3	8	0	1	20
131	7	6	4	0	4	17
132	3	6	4	4	3	20
133	12	4	4	4	6	30
134	10	9	8	0	5	32
135	6	3	5	4	4	22
136	6	5	2	0	1	14
137	10	6	0	9	6	31
138	10	3	7	1	4	25
139	3	2	11	4	0	20
140	0	4	7	0	4	15
141	12	0	12	4	12	40
142	6	3	5	12	0	26
143	0	6	10	11	1	28
144	10	3	11	6	7	37
145	3	6	11	1	11	32
146	10	3	10	4	4	31
147	6	3	8	6	5	28
148	10	6	8	1	6	31
149	10	4	12	4	8	38
150	3	3	0	10	3	19
151	10	3	7	8	2	30
152	6	1	2	3	12	24
153	10	4	11	1	4	30
154	10	3	10	7	0	30
155	10	2	4	3	11	30
156	12	8	10	1	2	33
157	12	7	7	3	1	30
158	10	3	4	4	2	23
159	3	6	5	1	2	17
160	10	3	7	3	4	27
161	10	3	7	1	3	24
162	0	2	8	2	0	12
163	10	4	11	1	2	28
164	10	3	10	4	0	27

Record No.	A[1]	B[2]	C[3]	D[4]	X[5]	S
165	12	4	0	6	4	26
166	6	4	8	4	4	26
167	10	4	8	3	2	27
168	10	4	0	7	4	25
169	3	4	11	3	5	26
170	3	1	4	3	11	22
171	7	0	0	11	5	23

(1) Original series minus 35 and remainder divided by 3

(2) " " " 30 " " " " 4

(3) " " " 120 " " " " 6

(4) " " " 238 " " " " 5

(5) " " " 95 " " " " 3

After the data has been coded and listed as on the preceding page* the values are then punched on punch cards, the record number also being punched for the purpose of identification. There are forty-five columns on a punch card. In our example the record number was punched in columns 31-2-3; A in columns 34-5; B in 36-7; C in 38-9; D in 40-1; X in 42-3; and S in 44-5. One card is used for each line--for each observation, that is.

After the cards have been punched, they should be summed: The sum of A, of B, of C, of D, of X and of S should be obtained; also the number of cards should be counted. The results should be recorded in some such form as follows:

Form 1

Sums and Means of the Variables.

# Items	A	B	C	D	X	S
171	1302	705	1149	680	774	4610
Means:	7.6140	4.1228	6.7193	3.9766	4.5263	26.9591

The use of the Check Sum first becomes apparent here: Evidently the sum of the sums of A, B, C, D, & X, should equal the Sum of S; which is the case. The same is true of the means (averages). This checks the first additions used in building up the check sum itself; it also checks the accuracy of the puncing; and also of the division in securing the averages. The values filled into the form above are for our example.

---

*It is sometimes feasible to do the coding by punching the orignal values upon the punch cards. Then sort the cards on the variable to be coded; group the arrayed values into the determined upon classes and gang punch each group in a new column with the assigned class value--such as 0, 8, &c.. The check sum for the individual record then can be prepared by showing each card separately in the tabulator, adding across, and subsequently punching upon the card, after which the procedure is as given above.

The next step is to sort the cards upon the first variable, A, the form for recording the tabulation is given below:

Form 2.

Sorted on:

Sorted Variable Class Value (1)	No. of Items (2)	A		B		C		D		X		S	
		Sum (3)	Ext. (4)	Sum (7)	Ext. (8)	Sum (9)	Ext. (10)	Sum (11)	Ext. (12)	Sum (13)	Ext. (14)	Sum (15)	Ext. (16)
Sums													

The cards being sorted on the first variable, A, group them into packs--all the cards of the lowest value of A in the first pack, of the next value of A in the second pack &c. &c.. List in column (1)--Form 2--the value of A in the first pack. Tabulate this pack. On the first line in column (2) write the number of cards in the pack; in column (5) write the sum of the values of A in this first pack, in column (7) write the sum of the values of B in this first pack; in column (9) the sum of the values of C in this pack; in column (11), D; in column (13), X; in column (15) S. Take the second pack, list the value of A in this pack on the second line of the form; and list the corresponding sum values as for the first pack. Repeat until all packs have been so treated. When this is completed make the extensions for columns 6, 8, 10, 12, 14, & 16 as follows:

Multiply the values listed respectively in columns 5, 7, 9, 11, 13 & 15 on any line each times the value listed on the same line in column 1. List the products so obtained in columns 6, 8, 10, 12, 14, & 16 respectively. Do this for all lines.

Add columns 2, 6, 8, 10, 12, 14 & 16.

Take the cards and sort them again; this time on the second variable, B. Take a second sheet (Form 2; Sh.2); divide the sorted cards into packs, according to the values of B and list these values of successive packs in column 1. Tabulate, list and extend in a manner exactly similar to that when the cards were sorted on A, except no figures need appear in the two A columns - 5 & 6.

Sort on C and proceed as for A & B on a new sheet (Form 2, Sh 3). No figures need appear in the A, or B columns: columns 5, 6, 7, & 8.

Sort on D and repeat (Form 2; Sh 4): No figures in the A, B, or C columns: Columns 5 to 10 inclusive.

Sort on X and repeat (Form 2, Sh 5). No figures in the A, B, C, or D columns, Columns 5 to 12 inclusive.

(Note: The reason that an increasing number of columns can be omitted is that to make the extensions and sum them would give figures already computed: Thus if we sort on C and extend its values times D, adding the extensions, we arrive at the same figure as if we had sorted on D and extended its values times C.) In case difficulty is encountered in making the figures check to the check sum -- explained later in connection with Form 3--it may be advisable to make the extensions here directed to be omitted, for the sake of comparisons and to help locate the errors.)

Following are the tabulations of the five sortings made in performing the above steps for our example. Note that in each case a check is afforded by adding up column 2. This should add to the total number of cards in the problem as shown by the data on form 1. A further check may be afforded by adding the sum columns for each variable--columns 5, 7, 9, 11, 13 & 15. These should on every sheet add to the same corresponding figures given on form 1.

Sorted on A

Form 2 Sheet #1.

Sorted Variable	Number	A		B		C		D		X		S	
Class Value	Items	Sum	Extension	Sum	Extension	Sum	Extension	Sum	Extension	Sum	Extension	Sum	Extension
(1)	(2)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
0	12	0	0	46	0	67	0	43	0	34	0	190	0
3	30	90	270	134	402	194	582	136	408	124	372	678	2034
6	28	168	1008	107	642	162	972	119	714	133	798	689	4134
7	6	42	294	25	175	29	203	38	266	31	217	165	1155
10	69	690	6900	270	2700	522	5220	236	2360	329	3290	2047	20470
12	26	312	3744	123	1476	175	2100	108	1296	123	1476	841	10092
	(171)		12216	(705)	5395	(1149)	9077	(680)	5044	(774)	6153	(4610)	37885

Sorted on B

Form 2 Sheet #2.

Class Value	Items	A Sum	A Extension	B Sum	B Extension	C Sum	C Extension	D Sum	D Extension	X Sum	X Extension	S Sum	S Extension
0	5			0	0	24	0	29	0	31	0	122	0
1	13			13	13	70	70	50	50	73	73	284	284
2	12			24	48	68	136	74	148	57	114	295	590
3	45			135	405	353	1059	169	507	228	684	1273	3819
4	30			120	480	180	720	109	436	119	476	777	3108
5	19			95	475	121	605	59	295	66	330	468	2340
6	31			186	1116	209	1254	144	864	123	738	888	5328
7	6			42	294	42	294	7	49	17	119	155	1085
8	4			32	256	37	296	17	136	18	144	137	1096
9	4			36	324	26	234	8	72	20	180	121	1089
10	1			10	100	11	110	3	30	11	110	38	380
12	1			12	144	8	96	11	132	11	132	52	624
	(171)				3655	(1149)	4874	(680)	2719	(774)	3100	(4610)	19743

Sorted on C

Form 2 Sheet #3.

Class Value	Items	A Sum	A Extension	B Sum	B Extension	C Sum	C Extension	D Sum	D Extension	X Sum	X Extension	S Sum	S Extension
0	21					0	0	166	0	89	0	454	0
1	3					3	3	22	22	13	13	65	65
2	7					14	28	29	58	30	60	148	296
4	13					52	208	44	176	79	316	333	1332
5	25					125	625	57	285	82	410	563	2815
7	19					133	931	39	273	90	630	499	3493
8	26					208	1664	86	688	127	1016	748	5984
10	21					210	2100	93	930	68	680	627	6270
11	28					308	3388	109	1199	141	1551	894	9834
12	8					96	1152	35	420	55	660	279	3348
	(171)						10099	(680)	4051	(774)	5336	(4610)	33437

Sorted on D

Form 2 - Sheet 14

Sorted Variable: Class Value	Number: Items	A Sum	A Extension	B Sum	B Extension	C Sum	C Extension	D Sum	D Extension	X Sum	X Extension	Y Sum	Y Extension
(1)	(2)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
0	28							0	0	136	0	706	0
1	29							29	29	128	128	737	737
2	6							12	24	31	62	154	332
3	30							90	270	130	390	761	1083
4	24							96	384	116	464	646	2584
5	3							15	75	7	35	62	310
6	10							60	360	43	258	376	1674
7	12							84	588	40	280	297	2079
8	8							64	512	31	248	238	1904
9	3							27	243	14	126	97	873
10	4							40	400	16	160	105	1050
11	5							55	605	41	451	196	2156
12	9							108	1296	41	472	201	2517
	(171)								4786	(774)	3094	(4510)	19694

Sorted on X

Form 2 Sheet 15

Class Value	Items	A Sum	A Extension	B Sum	B Extension	C Sum	C Extension	D Sum	D Extension	X Sum	X Extension	Y Sum	Y Extension
0	24									0	0	525	0
1	14									14	14	358	358
2	28									56	112	691	1382
3	12									36	108	278	834
4	25									100	400	646	2584
5	14									70	350	373	1865
6	12									72	432	333	1998
7	7									49	343	201	1407
8	5									40	320	139	1112
9	2									18	162	65	585
10	4									40	400	134	1340
11	9									99	1089	296	3256
12	15									180	2160	571	6852
	(171)										5890	(4510)	23573

The next step is to transcribe figures from the sheets (1 to 5) of Form 2, to another sheet of the form shown on the following page (Form 3).

On line A-1 (Form 3) column A list the figure taken from Form 2; Sh. 1 Col. 6, last (total) line. The other figures on line A-1 are taken from the same sheet (Form 2; Sh. 1) last line, columns 8, 10, 12, 14 & 16. The figures filled into form 3 apply to our example, so they may be traced through the various forms.

The data for line B-1 comes from Form 2, Sh. 2, last line, columns 8, 10, 12, 14, & 16. (This was the sheet used when the cards were sorted on B.)

The data for line C-1 comes from Form 2, Sh. 3, last line, columns 10, 12, 14, & 16. (This was the sheet used when the cards were sorted on C.)

The data for line D-1 comes from Form 2, Sh. 4, last line, columns 12, 14, & 16. (This was the sheet used when the cards were sorted on D.)

The data for line X-1 comes from Form 2, Sh. 5, last line, columns 14, & 16. (This was the sheet used when the cards were sorted on X) We have now filled in all that has to go on the lines ending in "1" (i.e.: A-1, B-1 &c &c) on Form 3.

To obtain the figures to go on the lines ending in "2" on Form 3, we make computations from the data on Form 1. The sum of the variable A ([illegible] in our example) is put into a computing machine and multiplied successively by the mean of A, of B, of C &c. &c.

Form 3
Transcription of Sums of Extensions.

Line #	A	B	C	D	X	S
A-1	12216.0	5395.0	9077.0	5044.0	6153.0	37885.0
A-2	9913.4	5367.9	8748.5	5177.5	5893.2	35100.5
A-3	2302.6	27.1	328.5	-133.5	259.8	2784.5
B-1		3655.0	4374.0	2719.0	3100.0	19743.0
B-2		2906.6	4737.1	2803.5	3191.0	19006.1
B-3		748.4	136.9	- 84.5	-91.0	736.9
C-1			10099.0	4051.0	5336.0	33437.0
C-2			7720.5	4500.1	5200.7	30975.9
C-3			2378.5	-518.1	135.3	2461.1
D-1				4786.0	3094.0	19694.0
D-2				2704.1	3077.9	18332.1
D-3				2081.9	16.1	1361.9
X-1					5890.0	23573.0
X-2					3503.4	20806.2
X-3					2386.6	2706.8

V

The products so obtained are listed respectively in Columns A, B, C &c., on Form 3, line A-2.

Next the sum of the second Variable, (B in our example; or 705) is put into the computing machine and multiplied successively by the mean of B, of C, of D &c. &c.. The products so obtained are listed respectively in Columns B, C, D, &c. of Form 3, line B-2.

Next the sum of the third Variable, (C in our example; or 1149) is put into the computing machine and multiplied successively by the mean of C, of D &c. &c.. The products so obtained are listed respectively in Columns C, D, &c. of Form 3, line C-2.

In a similar manner the computations are made for the other lines ending in "2", Form 3.

Note:- In practice it is most convenient to prepare Form 3 on a sheet of paper which also carries Form 1 at the top. The figures for making the extension for lines of designation ending in "2" are then before the operator.

Every item on a line ending in "2" is now subtracted from the figure directly above it on a line ending in "1." (Note: naturally should the minuend be greater than the subtrahend the difference will be a negative value.) These differences are listed on the lines of designation ending in "3". These lines are then transcribed to Form 4.

The differences which have just been secured are the product moments and squared standard deviations (times N); and are the necessary data for any solution of multiple, and partial correlation coefficients, or gross and net regression coefficients. The usual solution may be found in Yule: "Introduction to Statistical Method." The solution given in the following, however, is a "Least Square" method, first conceived of and developed by Messrs. Tolley and Ezekiel of the Bureau

of Agricultural Economics. An article describing the theory of the method is published by them in the Journal of the Am. Stat. Assoc., for December, 1923.

Note:- In case it is decided to make the extensions by hand rather than to use punch cards and tabulating machines--frequently the case when short series, such as time series, are being analyzed,--a multi-columnar form should be used. In the six left-hand columns list the data, column headings would be A, B, C, D, X, & S. The remaining columns should be headed: A^2, AB, AC, AD, AX, AS, B^2, BC, BD, BX, BS, C^2, CD, CX, CS; D^2, DX, DS; X^2, XS. In the A^2 column write the squares of the values in the A Column. In the AB column write the products of the A items times their corresponding B items, &c &c. When all columns have been extended, add them, listing the totals below, in their respective columns.

Find the means (averages) of the A, B, C, D, X, & S columns. Multiply the sum of the A column times each of the means of the A, B, C, D, X, & S columns and write the products below the sums of the A^2, AB, AC, AD, AX, & AS columns respectively. Multiply the sum of the B column times the means of the B, C, D, X, & S columns and write the products below the sums of the B^2, BC, BD, BX, & BS, columns. In a similar manner extend the sum of the C column times the means of C, D, X, & S, and inscribe the products in the C^2, CD, CX, & CS columns. Also the sums of the D, & X columns. It is not necessary to multiply the sum of the S column times anything.

Now subtract the last values listed in the A^2 column and in columns to the right thereof from the figures just above them. If the minuend should be greater than the subtrahend the difference would naturally be a negative value. These differences are now to be transferred to a new sheet of the arrangement shown in form 4. The differences in the columns commencing with an "A" in their designation are transferred to the first line of form 4, designated as line A-3. The differences in the columns commencing with a "B" in their designation (this of course includes the B^2 col.) are transferred to line B-3 of form 4. The remaining differences are transferred in a similar manner. The so arranged differences constitute the Normal Equations of a least square solution for the value of the net regressions of A, B, C, & D, on X.

At this point the use of the Check Sum (Col. S) as a check on the work to this point may be shown: On Line A-1 (Form 3) the sum of the items in columns A, B, C, D & X should equal the figure in Column S (Line A-1), thus checking the extension and addition of all the figures used in connection with deriving these values.

	Line	Column
The Sum of the following:	A-1	B
	B-1	B
	"	C
	"	D
	"	X
Should check to:		
	B-1	S
The Sum of the following:	A-1	C
	B-1	C
	C-1	C
	"	D
	"	X
Should check to:		
	C-1	S
The Sum of the following:		
	A-1	D
	B-1	D
	C-1	D
	D-1	D
	"	X
Should check to:		
	D-1	S
The sum of the following:		
	A-1	X
	B-1	X
	C-1	X
	D-1	X
	X-1	X
Should check to:		
	X-1	S.

By substituting "2" for "1", and also "3" for "1" in the above schedule, a further check may be secured. It is essential that the values should all check, before the work is carried to a further stage

(Form 4)

NORMAL EQUATIONS

Line#		A	B	C	D	X	S	PM
A-3		2302.6	27.1	328.5	-133.5	259.8	2784.5	
B-3			748.4	136.9	-34.5	-91.0	736.9	
C-3				2378.5	-518.1	135.3	2461.1	
D-3					2081.9	16.1	1361.9	
X-3						2386.6	2706.8	

SOLUTION

Line#		A	B	C	D	X	S	PM
1		2302.6	27.1	328.5	-133.5	259.9	2784.5	
2		-1.0000	-.0118	-.1427	.0580	-.1128	-1.2093	
3			748.4	136.9	-34.5	-91.0	736.9	
4			-.3	-3.9	1.6	-3.1	-32.8	
5			748.1	133.0	-82.9	-94.1	704.1	
6			-1.0000	-.1778	.1108	.1258	-.9412	
7				2378.5	-518.1	135.3	2461.1	
8				-46.9	19.0	-37.1	-397.2	
9				-23.7	14.8	16.7	-125.2	
10				2307.9	-484.3	114.9	1938.7	
11				-1.0000	.2098	-.0498	-.8400	
12					2081.9	16.1	1361.9	
13					-7.7	15.1	161.4	
14					-9.2	-10.4	78.0	
15					-101.6	24.1	406.8	
16					1963.4	44.9	2008.1	
17					-1.0000	-.0229	-1.0228	
18					D	: .0229	x 16.1 :	.4
19				C: .0498	.0048	: .0546	x 135.3 :	7.[illegible]
20		B:	-.1258	-.0097	.0025	:-.1330	x -91.0 :	12.[illegible]
21	A	.1128	.0016	-.0078	.0013	: .1079	x 259.8 :	28.0
22							P.M.:	43.0
23							Sq. Root :	6.94
24		243.45	3.60	17.9	-3.05	: 250.79		
25		Square root of 2386.6 (See Line X-3; Col.X)					:	48.8

R equals 6.94 ÷ 43.8 or .142

On Form 4, Lines A-3, B-3, C-3, and D-3, are the data comprising the normal equations which are to be solved. The method of solution is given on lines 1 to 25, as described below.*

On Line 1 write the first normal equation, i.e., copy Line A-3. Now, divide every item on Line 1 by the first item of Line 1, reverse the algebraic signs and list the quotients on Line 2. In our example, we divide by 2302.6. The algebraic sum of the items on Line 2 Columns A, B, C, D, and X, should equal the quotient appearing in Column S on the same line. This will not always check to the last digit, owing to the dropping of places in the division. Now, draw a line under the figures just written in. On Line 3 copy in the second normal equation, that is, copy Line B-3. Now put the figure on Line 2, Column B (i.e. -.0118) into the multiplying machine and multiply it consecutively by the items on Line 1 in Columns B, C, D, X, and S, listing the products in the respective columns on Line 4. In our example, we multiply -.0118 by 27.1 by 323.5 by .133.5 by 259.9 and by 2734.5, giving as quotients the items appearing on Line 4, i.e., giving -.3, -3.9; 1.6; -3.1; -32.8. Now add the items on Line 4 to the items immediately above on Line 3, giving Line 5. Careful attention must be given to the algebraic signs. Now, divide the figures on Line 5 by the first figure on Line 5, and reverse the algebraic signs, listing the quotients on Line 6. In our example, we divide by 748.1. The algebraic sum of the first four items on Line 5 should check to the last item of the line, 704.1, in the S column. In like manner, the algebraic sum of the first four items on

---

*This is the "Doolittle Method" - See Oscar S. Adams - "Geodesy - Application of the Theory of Least Squares to the Adjustment of Triangulation." - 1915. Special publication #28, Geodetic Survey.

Line 6 should check to the last item on Line 6, or -.9412. Now copy down the third normal equation on Line 7; i.e., copy Line C-3. Put the number in the C column on the second line, -.1427 into the multiplying machine and multiply it consecutively by the items in the C, D, X, and S columns of Line 1, listing the products in the corresponding columns on Line 8, giving careful attention to the algebraic signs. Next, put the item in the C column of Line 6, -.1778, into the multiplying machine and multiply it consecutively by the C, D, X, and S column figures on Line 5, listing the products in their respective columns on Line 9, giving careful attention to algebraic signs. Now, add together for each column the values in Lines 7, 8, and 9, giving Line 10. The first items of this line should check to the last, similar to the case for Lines 5 and 1. Divide each of the items of this line by the first item in the line; that is, divide by 2307.9, reverse the signs and list the quotients on Line 11. In a manner similar to Line 6, the first items on this line should check to the last when added together. Draw another line. On Line 12, write the fourth normal equation, that is copy Line D-3. Put into the multiplying machine the value on Line 2, Column D, and multiply it consecutively by the D, X, and S column values of Line 1, listing the products on Line 13 in their respective columns, giving careful attention to the algebraic signs. Next, place the value on Line 6 column D into the multiplying machine, .1108, and multiply consecutively by the values on Line 5, columns D, X, and S, listing the products in their respective columns on Line 14, giving careful attention to the algebraic signs. Next, put the figure on Line 11, column D, .2093, into the multiplying machine and multiply consecutively by the values on Line 10 columns D, X, and S, listing the products in their

respective columns on Line 15, giving careful attention to algebraic signs. Next, add for columns D, X, and S the values of Lines 12, 13, 14, and 15, listing the algebraic sums on Line 16. Careful attention should be given to algebraic signs. The algebraic sum of the first items on Line 16 should check to the last item, as was the case for Lines 10, 5, and 1. Divide the items on Line 16 by the first item on Line 16, 1963.4, reverse the algebraic signs, and list the quotations on Line 17. The first items of Line 17 should check to the last item of Line 17, as was the case for Lines 11, 6, and 2. We have now finished the "forward" solution for the 4 normal equations, and by this time the reader presumably understands the method so that the extension of the solution to a greater number of variables will be a comparatively simple matter. We are now ready for the back solution, which is given on Lines 18, et seq.

On Line 18, column X write the value on Line 17 column X; reverse the sign. This value is the net regression coefficient of the variable D on X. Next in Column S Lines 18 to 21, inclusive, list in inverse order the values in column X, Lines A-3, B-3, C-3, D-3. Next on Line 19 Column C, write the value on Line 11 Column X, reversing the sign. On Line 20, column B, write the value on Line 6, column X, reversing the sign. On Line 21, column A, write the value on Line 2 column X, reversing the sign. Put the value on Line 18, column X, into the multiplying machine and multiply it times the value on the same line in column S, and list the product in Column P (Product Moment). Then multiply it times the values in column D, Lines 11, 6 and 2, listing the products respectively on Lines 19, 20 and 21 in column D. Multiply it also times the value in column D on Line A-3, listing the product in Column D on Line 21.

Next, add together the values on Line 19 in columns C and D, listing their sum in column X, Line 19. This last sum is the net regression coefficient of the variable C on X.

Put the net regression coefficient of the variable C on X into the multiplying machine, and, having a care for algebraic signs, multiply it times the value listed beside it in column S, writing the product on the same line in column PM. Then, multiply it by the values in column C on Lines 6, 2 and A-3, writing the products on Lines 20, 21 and 24, respectively of the same column. Now, add the values on Line 20, columns B, C and D together writing the sum in column X on the same line, having a particular regard to algebraic signs. This last sum written on Line 20 in column X is the net regression coefficient of the variable B on X. Place it in the multiplying machine, and, having a care to algebraic signs, multiply it times the value listed beside it in column S, writing the product on the same line in column PM. Then, multiply it by the values in column B on Lines 2 and A-3, listing the products respectively on Lines 21, and 24 of the same column. Now, add the values on Line 21, and in columns A, B, C, and D, writing the sum in Column X on the same line. This sum is the net regression coefficient of A on X. Place it in the multiplying machine and multiply it times the value listed beside it in column S, having a care for algebraic signs, and list the product in column PM on the same line. Also, multiply it times the value on Line A-3, column A, listing the product in the same column on Line 24. Now, the values on Line 24 in columns A, B, C, D added together algebraically, should equal the value listed on Line A-3, column X, which serves as a check upon the derivation of other net regression coefficients than D on X. There is a difference

of 9 between the two values in our example. A greater accuracy may be secured carrying the arithmetic to a greater number of places throughout the entire solution. It was deemed expedient to make the example as simple as possible.

We have now secured the net regression coefficients, which are essential to the forming of the regression equation for predicting or estimating the values of S. To ascertain to how great a degree these predictions conform to the actual values, it is necessary to obtain some measure of agreement between them, the predicted and the actual. This measure of agreement is the coefficient of multiple correlation, R, defined on Page 1. To secure this coefficient of multiple correlation, add the values in the PM column, listing the sum on Line 22. Next, secure the square root of this sum, given on Line 23. Finally, secure the square root of the value listed on Line X-3, column X, listing this in the PM column on Line 25. This value divided into the value immediately above it on Line 23, gives the coefficient of multiple correlation.

There are certain aids and other checks in the solution which can be applied to help in locating errors. The diagonal terms of the normal equations (2302.6, 743.4, 2373.5 &c) are always positive in sign. In making the solution the figures listed immediately below these figures (to be added to them in the course of the solution) are always negative in sign. The sums appearing above the -1.00000 terms are always positive, (i.e. 743.1, 2307.9, 1963.4).

Accuracy is increased if comparatively small diagonal terms are avoided. (This can be controlled by controlling the original coding.)

The Product Moment, Line 22, is always positive in sign.

THE REGRESSION EQUATION.

The final step in the arithmetic is to write the "regression," or "prediciting" or "estimating" equation as it is variously called. First write down the regression coefficient of A on X; in our example this is .1079. Beside this value write what was done algebraically in simplifying the A series: this will be of the form, $\frac{A - 35}{3}$. (See page 6, note (1)). Then further write the subtraction of the average of the A series as taken from form 1, enclosing all in parentheses. The algebr c form so far will look like this: $.1079 \left(\frac{A - 35}{3} - 7.6140 \right)$

In an exactly similar manner treat the B, C, D, & X series. (Disregard the S series) There will be no regression <u>coefficient</u> for the X series. The algebraic sum of the A, B, C, & D expressions should be equated to the X expression as follows:

$$+ .1079\left(\frac{A-35}{3} - 7.614\right) - .1330\left(\frac{B-30}{4} - 4.1226\right) + .0546\left(\frac{C-120}{6} - 6.7193\right) \ldots\ldots$$

$$+ .0229\left(\frac{D-238}{5} - 3.9766\right) = \frac{X - 95}{3} - 4.5263$$

This is the "raw" regression equation. It is only necessary now to evaluate for X, involving only elementary algebra, to put the equation in its most useful form:

$$X = .1079\,A - .0998\,B + .0273\,C + .0137\,D + 93.0552$$

This is the "predicting" equation mentioned in the definition of R in the note on page 1.